

Automatic reaction network generation using chemo-informatics

Nick M. Vandewiele, Kevin M. Van Geem, Marie-Françoise Reyniers, Guy B. Marin

Laboratory for Chemical Technology, Krijgslaan 281 (S5), 9000 Gent, Belgium

Large-scale detailed kinetic models find increasing use in the modeling of combustion processes, atmospheric chemistry, soot formation, and other areas of industrial or environmental interest. The enormous amount of reaction possibilities in these chemical processes lead to complex reaction networks consisting of hundreds of unique species and thousands of unique reactions. The manual construction of these networks is therefore in most cases impossible, and in the few cases it is possible it remains tedious and error-prone. Hence, automatic generation of a chemical reaction network is particularly useful for these complex reactive processes such as pyrolysis, gasification, oxidation, combustion, etc.

One of the most important aspect of computer tools that enable the automated execution of chemical transformations is the representation of chemical species. Graphs, consisting of vertices and edges, are an excellent mathematical abstraction of chemical species and were the starting basis for the majority of the tools. Moreover, the rise of chemo-informatics, i.e. the use of informatics methods to solve chemical problems, at the intersection between chemistry and computer science, responded to the needs to chemical community to store and analyze the ever-growing amount of chemical data available. Chemo-informatics unlocked a broad number of algorithms emanating from graph theory to chemists and chemical engineers, leading among others to unique molecule identifiers (SMILES [1], InChI [2]) and powerful substructure matching algorithms. However, these advanced representations and algorithms are only rarely picked up by researchers in the field of reaction engineering. Therefore we have developed a new automatic reaction network generation tool entitled 'Genesys' that takes full advantage of the recent advances made in the field of chemo-informatics. In Genesys graph theory algorithms are applied originating from open-source chemo-informatics libraries such as the *Chemistry Development Kit* [3]. Genesys generates a reaction network consisting of elementary reactions based on a number of user-defined reaction families. Each reaction family consists of a recipe containing the elementary actions to convert reactants into product species, a description of the required sub-molecular pattern inside candidate molecules and a set of constraints, structural features of species that prevent candidate molecules from undergoing a specific reaction family.

Sub-molecular patterns are unambiguously defined using the SMARTS (SMILES Arbitrary Target Specification) [4] language. The identification of these sub-molecular patterns allows to implement Benson's group additivity scheme [5] for the estimation of thermodynamic properties of chemical species in a straight forward way via a combination of group-additive values, ring-strain and non-nearest neighbor corrections.

The application of the reaction network generation program and its general applicability to both catalytic and non-catalytic processes will be discussed. For illustrative purposes

different reaction networks have been automatically generated for the pyrolysis of ethane/toluene. The chemical knowledge of the user is reflected by the user-defined set of reaction families and allows to control reaction network size and generation time. Post-processing options such as species and reactions visualization, compatibility with reactor modeling tools such as Chemkin, and tools to analyze the generated reaction mechanism will be discussed.

References

- [1]. Weininger, D., SMILES, A CHEMICAL LANGUAGE AND INFORMATION-SYSTEM .1. INTRODUCTION TO METHODOLOGY AND ENCODING RULES. *Journal of Chemical Information and Computer Sciences* **1988**, 28, (1), 31-36.
- [2]. Heller, S. R.; Stein, S. E.; Tchekhovskoi, D. V., InChI: Open access/open source and the IUPAC international chemical identifier. *Abstracts of Papers of the American Chemical Society* **2005**, 230, 60-CINF.
- [3]. Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E., The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics. *J Chem Inf Comput Sci* **2003**, 43, (2), 493 - 500.
- [4]. Daylight Chemical Information Systems, I. Daylight Theory Manual.
<http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>
- [5]. Cohen, N.; Benson, S. W.; Patai, S.; Rappoport, Z., The thermochemistry of alkanes and cycloalkanes. In *The chemistry of alkanes and cycloalkanes*, Wiley: Chichester, 1992; pp 215-288.